ReputationNet: Reputation-Based Service Recommendation for e-Science

Jinhui Yao, Wei Tan, *Senior Member, IEEE*, Surya Nepal, Shiping Chen, *Senior Member, IEEE*, Jia Zhang, David De Roure, and Carole Goble

Abstract—In the paradigm of service oriented science, scientific computing applications and data are all wrapped as web accessible services. Scientific workflows further integrate these services to answer complex research questions. However, our earlier study conducted on myExperiment has revealed that although the sharing of service-based capabilities opens a gateway to resource reuse, in practice, the degree of reuse is very low. This finding has motivated us to propose ServiceMap to provide navigation facility through the network of services to facilitate the design and development of scientific workflows. This paper proposes ReputationNet as an enhancement of ServiceMap, to incorporate the often-ignored reputation aspects of services/workflows and their publishers, in order to offer better service and workflow recommendations. We have developed a novel model to reflect the reputation of e-Science services/workflows, and developed heuristic algorithms to provide service recommendations based on reputations. Experiments on myExperiment have illustrated a strong positive correlation (with Pearson correlation coefficient 0.82) between the reputation scores computed and the actual performance (i.e. usage frequency) of the services/workflows, which demonstrates the effectiveness of our approach.

Index Terms—Service oriented architecture, scientific workflow, reputation, service composition

1 Introduction

Service Oriented Architecture (SOA) enables flexible and dynamic collaborations among different service providers. A software service can either be used in a standalone manner or be composed with others to form a value-added service [35]. In this e-Science paradigm, many scientific disciplines such as physics and biology have embraced SOA to integrate their heterogeneous data and computation resources. A scientific workflow precisely defines a multistep procedure to seamlessly integrate and streamline local and remote heterogeneous computational and data resources to answer complicated research questions [42]. For two examples, in astronomy, scientists run workflows to generate science-grade mosaics of the sky [4] and examine the structure of galaxies [43]. In bioinformatics, workflows help reason rare illness by processing a large amount of bimolecular data and understanding their statistical nature [33], [34].

Scientific investigation has entered the age of *data-intensive science*, or *e-Science* [41]. In e-Science, as scientists create and use more and more services (such as caBIG [38] and KEGG [21]) and workflows, they have an increasing inclination to publicize these workflows to disseminate their experimental results and obtain credit. This phenomenon has a far reaching impact, i.e., a scientific workflow can be wrapped up as a universal service that can either be later reused directly or adapted and repurposed. As a matter of fact, several domain-specific online workflow repositories have evolved in recent years. Among them, the largest collection is the UK-based myExperiment [14] project. myExperiment is a social web site for researchers sharing and using scientific workflows and other research objects, with a collection of than 2,800 life-science workflows as of May 2013. The advent of these online repositories makes it possible to assess the state-of-art of scientific workflows and promote the reuse of the best practices embedded therein. However, our earlier study [39] conducted on myExperiment indicated that although the sharing of service-based capabilities opens a gateway to resource reuse, in practice, such reuse rate is low. In response, the ServiceMap framework [42] was developed to use association rule mining and matrix-based searching algorithms to help domain scientists better understand various usage patterns of the existing services; and provide a system level support to recommend possible service candidates and their compositions.

ServiceMap recommends a collection of services that are frequently used together or a service chain serving a specific purpose. Its performance is based on network statistics and matrix calculation, without taking into account the social factor [26], [18] of workflow usage. However, the e-Science ecosystem is not a pure software system. Instead, it consists of not only services and workflows, but also corresponding creators (authors) and users. The authors create services, and may use services created by others to build

- J. Yao is with Palo Alto Research Center (PARC), Webster, NY. E-mail: jinhui.yao@xerox.com.
- W. Tan is with IBM T.J. Watson Research Center, New York, NY. E-mail: wtan@us.ibm.com.
- S. Nepal and S. Chen are with CSIRO Computational Informatics (CCI), Sydney, NSW, Australia. E-mail: {surya.nepal, shiping.chen}@csiro.au.
- J. Zhang is with Carnegie Mellon University-Silicon Valley, San Francisco, CA. E-mail: jia.zhang@sv.cmu.edu.
- D. D. Roure is with Oxford e-Research Centre, Oxford University, Oxford, United Kingdom. E-mail: david.deroure@oerc.ox.ac.uk.
- C. Goble is with the School of Computer Science, University of Manchester, United Kingdom. E-mail: carole.goble@manchester.ac.uk.

Manuscript received 1 Dec. 2013; revised 15 May 2014; accepted 22 July 2014. Date of publication 19 Oct. 2014; date of current version 12 June 2015.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TSC.2014.2364029

value-added workflows. They may also jointly create services/workflows, provide help and support to one another on how to use services/workflows, and form a social network where they rate services/workflows and their creators. We expect that these social network interactions can bring additional information to make service recommendation more personalized and accurate. In this paper, we use trust on service/workflow and its publisher, as perceived by consumers and other publishers, for selecting the best option. The reputation mechanisms have been widely considered as an effective approach [24], [25] to evaluate the extent to which one trusts another. Therefore, we extend the ServiceMap framework and generate the ReputationNet by incorporating the trust of both services/workflows and their publishers into the service network, to reinforce the capability of ServiceMap in terms of service and workflow recommendations.

Our contributions are summarized as follows. First, we have developed a novel modeling approach to evaluate and present the reputation aspect of services, workflows and workflow authors, in the context of e-Science. Second, we have developed heuristic algorithms to derive a variety of reputation scores and provide service recommendations based on reputations. We have conducted a range of experiments using the workflows on myExperiment to evaluate our ReputationNet. Our experimental results have displayed a strong positive correlation (with Pearson correlation coefficient 0.82) between the reputation scores computed and the actual performance of the service/workflow (i.e. usage frequency). Strong correlations are also evident between reputation scores computed from ReputationNet and other service subjects (e.g., service compositions and service association groups). In a word, ReputationNet offers a good measurement of reputation of services and workflows.

The rest of the paper is organized as follows. Section 2 presents the motivation of our work as well as some background of ServiceMap. Section 3 shows an overview of our reputation-based approach to leverage and complement ServiceMap, and describes the models developed. Section 4 presents algorithms to utilize our modeling for recommendation. Section 5 presents the results of our evaluations and experiments. Section 6 discusses related work and section 7 concludes the paper.

2 Background and Research Motivation

Our ServiceMap [42] framework addresses two questions that domain scientists frequently ask when exploiting external Web services in building a scientific workflow:

Q1: *Given the services I plan to use, what are the other services often used together with them, by other scientists?*

Q2: *Given two or more services I want to use together, can I find an operation chain, which is already used by others, to connect them?*

In an attempt to provide the answers, ServiceMap was developed as a framework. First, all myExperiment workflows are downloaded. These workflows are then abstracted by removing non-Web services (i.e., WSDL-based and RESTful services), such as local beanshells, xml manipulating blocks, while maintaining the data flows between services. Afterwards, the abstract workflows are inputted into ServiceMap. ServiceMap consists of two disjoint networks (graphs): an undirected workflow-service network and a directed operation network. In the undirected workflow-service network (Fig. 1), nodes are either workflows or services and edges represent the inclusive relations between them—that is, the associations among the workflows and services. In the directed operation network (Fig. 2), nodes are operations in services, and a directed edge represents a data link between two operations in some workflow. More details regarding the myExperiment workflow set, how networks are built and analyzed can be found in [39], [42].

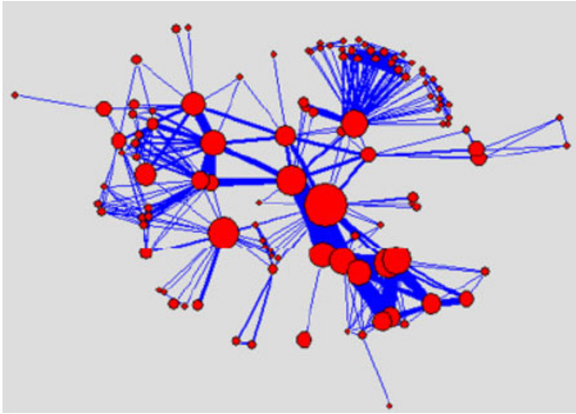


Fig. 1. Service association network (partial).

To answer Q1, we derive the frequent item sets and associations rules in the workflow-service network, and recommend relevant services in a given context (i.e., existing services in a workflow). To answer Q2, we have developed a relation-aware, cross-workflow search method to identify an operation chain which connects two services and is composed by fragments from individual workflows.

To utilize ServiceMap for service recommendations, intuitive enough, one may need to find which services have the strongest association with the target service and they can be recommended with higher confidence; and one may need to find which service composition is more likely to exceed the others when there are multiple reachable links between given services. However, this task is not trivial. An empirical study on the two networks in the ServiceMap has revealed very limited reusability of the services and the service compositions among the workflows. Thus, the likelihood of a service or a service composition being used in multiple workflows is low. This brings significant difficulties for service recommendation.

Two examples are shown in Fig. 3. The directed graph in Fig. 3a illustrates a case where three different paths (service compositions) are found linking service a (WSDbfetch.wsdl) and b (WSClustalW2.wsdl). The undirected graph in Fig. 3b shows a case that service c (WSWUBlast.wsdl) is found to have associations with four other services (i.e., the services used together in the same workflow). The two cases are derived by ServiceMap using real-world workflows stored on myExperiment. The numbers on the directed/undirected links indicate the number of occurrences of the links, e.g., three means three workflows have used such

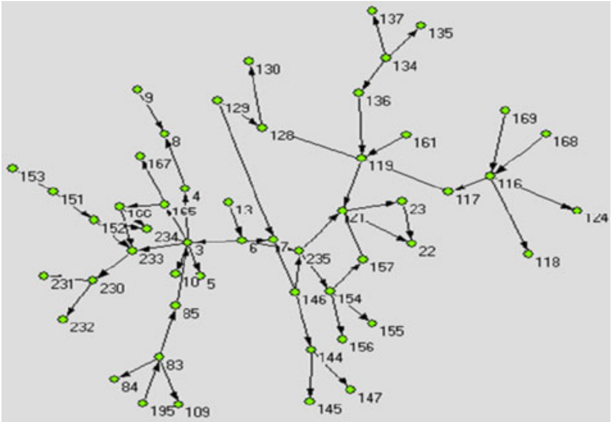
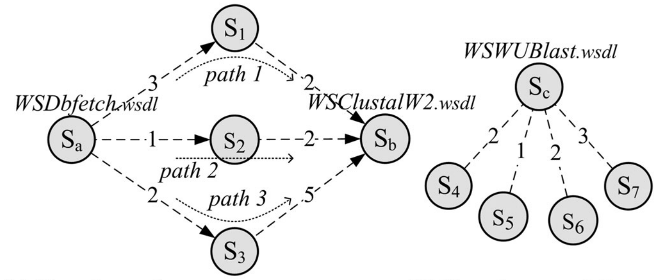


Fig. 2. Service operation workflow (partial).)

service composition. However, based solely on the numbers of occurrences to identify the most optimal path or association remains challenging. Considering the number of total workflows available (more than 2,800 as of May 2013), none of the paths or associations has the number of occurrences significantly exceeds those of others. Meanwhile, many other factors may have an impact on this. For instance, the credentials of the authors of a workflow may suggest the credibility of a particular service composition or association, e.g., if path_2 is the only one designed by an expert in the field, although it includes smaller number of occurrences of the links compared to the alternatives, path_2 may be considered more trustworthy. Another factor to consider may be the popularity, as the adoption of a workflow could be an indicator of its reliability and usefulness.

While the community of myExperiment is developing rapidly, the sparseness discovered in our empirical study and the lack of important factors [40], [42] have indicated the ineffectiveness of using simple frequency-based recommendation approaches. This has revealed the necessity to employ alternative methods to complement the network



(a) Choosing path (b) Choosing association

Fig. 3. Motivating examples.

analysis of ServiceMap, in order to deliver accurate and effective recommendations. This motivates the development of ReputationNet to consider the reputation aspects of workflows and their designers to reinforce the ServiceMap.

3 Reputation Based Service Recommendation

We define trust as the belief that a user holds regarding the intention and capability of a service/workflow to behave as expected. We use reputation as a mechanism of establishing the belief on a service and its publishers ability to deliver, through collective perception of the users/workflows that have interacted with the service in the past. This mechanism has been successfully applied in Internet marketplaces such as eBay and Amazon, as well as Web services using the concept of reputation [27]. The notion underpinning the reputation-based trust models is to capture consumers' perception of the consumed service and use it to evaluate the reputation of the service [8]. A social platform like myExperiment provides a rich collection of reputation data that can be analyzed.

Fig. 4 depicts the overall approach of our ReputationNet. We first download all the workflows from myExperiment to build the ServiceMap, which constructs both an undirected

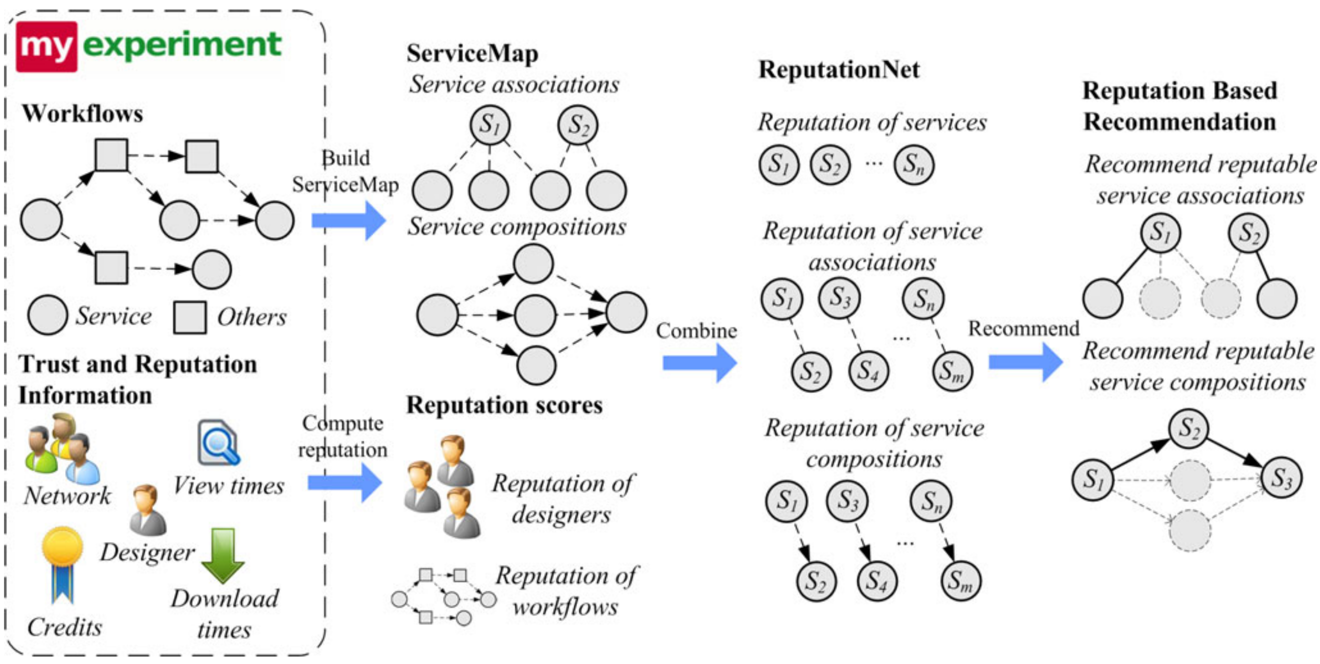


Fig. 4. The ReputationNet approach for service recommendation.

TABLE 1
Table of Symbols

w, W	Workflows and workflow set
s, S	Service and service set
d, D	Designers and designer set
$s^{\{\}}, s^{<>}$	Service composition and service association
p	Popularity
r, R	Individual reputation and aggregate reputation
n_v, n_d	Number of views and number of downloads
Rt	Ratings
Ctr	Contributions
α_{fd}, β_{fd}	Reputation and popularity fading factors

network capturing the associations between the services, and a directed network capturing the service compositions. The method of building the ServiceMap is reported in [42]. In the meanwhile, we download other reputation information about the workflows and their designers to compute their reputation scores. These information include the view/download times of each workflow, credits received by designers, the friends to whom each designer is connected (their uses will be elaborated in later sections). By combining the network structure of the services, the service associations as well as the service compositions with the reputation scores, we derive the ReputationNet. ReputationNet enables us, for instance, to identify the most reputable association (an extension to Q1) that exists between a given service and others related to it. Similarly, for two or more services, we can identify a path linking them that has the highest reputation (an extension to Q2).

3.1 The Model of Service Reputation

In our model, we view platforms like myExperiment as a giant workflow repository, whereas the members of the platform constantly propose and add new workflows into it. Each member can propose workflows and view or download the ones proposed by others. Moreover, new workflows can be developed based on existing ones, e.g., to combine two workflows to form a new one. Web services can be used as building blocks inside of workflows which may appear in the form of a single service or service compositions. More formally, we can model a repository (*Rep*) as a tuple

$$Rep = \left(D, W, S, S^{\{\}}, S^{<>}\right), \tag{1}$$

where $D = \{D_1, D_2, \ldots, D_d\}$ represents the set of registered members, the designers (e.g., scientists) who develop the workflows. Therefore, $W = \{W_1, W_2, \ldots, W_w\}$ is the set of the workflows proposed. Workflows are defined by the designer, that is $D_i \rightarrow^{define} W_j$ where $D_i \in D$ and $W_j \in W$.

Workflows may be viewed as sequences of activities which are to be carried out to accomplish a specific goal. Such activities may exhibit in a variety of forms, among them, there are invocations to certain local or remote services. Here the set $S = \{S_1, S_2, \ldots, S_n\}$ represents the services that have been used in the workflows defined, thus $S \in W$. As mentioned, in a workflow, multiple services may be used as a service compositions, a service composition is a collection of services which are executed according to a specific order, e.g., $S_1^{\{\}} = S_1 \rightarrow S_2 \rightarrow S_3$ represents a service composition where three services (i.e. S_1, S_2 and S_3) are executed in sequential order. $S^{\{\}} = \{S_1^{\{\}}, S_2^{\{\}}, \ldots, S_c^{\{\}}\}$ is the set of the service compositions that have been used, the same as services, $S^{\{\}} \in W$. The element $S^{<>} = \{S_1^{<>}, S_2^{<>}, \ldots, S_n^{<>}\}$ is the set of the associations among the services in S , which captures the instances of multiple services being used together.

The reputations of the proposed workflows are, thus, evaluated according to two main factors, namely, the reputations of the designers (R_{Di}) and the popularities of the proposed workflows (p^{Wi}). The reputation of workflow W_i proposed by designer D_i is thus computed using the following formula:

$$r_{Di}^{Wi} = f_r(R_{Di}, p^{Wi}), \tag{2}$$

where function f_r must be a monotonic increasing function, i.e., in numerical terms, $R_{D1} > R_{D2}$ and $p^{W1} > p^{W2}$ imply $r_{D1}^{W1} > r_{D2}^{W2}$. This modeling approach is based on the common empirical experience, whereby, the more reputable the designer is (i.e. a higher R_{Di}), and the more times the workflow has been used by others (i.e. a higher p^{Wi}), the more likely it is that the workflow is of a higher quality.

The popularity of a workflow (p^{Wi}) should reflect the degree of adoption by other users in the community. While it is difficult to establish exactly how many users have actually used a certain workflow, some indicating statistics are at our disposal, such as the number of times it is viewed (n_v^{Wi}), downloaded (n_d^{Wi}), as well as the ratings or feedbacks given by other users ($\{Rt^{Wi}\}$). The intuition behind this is that, the more frequently a workflow is accessed by users and the higher its user rating is, the more popular it is in the community. The popularity can be computed as follows:

$$p^{Wi} = f_p\left(n_v^{Wi}, n_d^{Wi}, \{Rt^{Wi}\}\right). \tag{3}$$

The computation of the author's reputation should take into account his/her contributions to the community as well as his/her popularity among the users. The social network platform may have certain mechanisms to reflect one's contributions. For example, in myExperiment each user has a credit score and an average rating. Moreover, the popularity of the workflows that the user has contributed in the past directly suggests his/her reputation as a credible designer. Similarly, a particular user's popularity in the community can be assessed according to his/her connections. The more reputable members one has in his/her connections, the more likely he/she is also a reputable member in the corresponding field.

To capture one's contributions to the community in terms of the quality of the workflows he/she has proposed, we here define the contribution of designer D_i as follows:

$$Ctr_{Di} = f_{Ctr}(C_{Di}, \{p_{Di}\}), \tag{4}$$

where C_{Di} refers to the credit score (or other forms of evaluation) provided by the social network platform, which indicates ones contributions in the workflows proposed by other members; $\{p_{Di}\}$ is the collection of the popularities of the workflows that user has proposed, which is an indication of the contributions one has made directly. The credit score may be computed in various ways, for instance, if the member is assumed to share all contributions of the workflows he has contributed, the popularities of those workflows can be used as the credit score.

Ones reputation in the community depends on both his contributions the recognition from the users; and the contributions made by his connections the recognition from other reputable designers. The reputation of a designer is defined as

$$R_{Di} = f_{Rd}\big(Ctr_{Di}, \{Ctr^{Di}_{cnt}\}\big), \tag{5}$$

where $\{Ctr^{Di}_{cnt}\}$ is the collection of the contributions of the members to whom D_i is connected. The underpinning intuition is that, if a designer is connected to many high contribution professionals in the community, then it is likely this designer is also professional and reputable. This score represents the fame as well as the recognition of the designer in the scientific community, and thus is a good indication of the quality of the designed workflows.

Each Workflow may contain one service or a collection of services, which, in turn, may form a service composition. Here, the term 'service composition' has a broader meaning, as in the scientific workflows services may comprise of other local non-service activities. Hence, in this context, service composition refers to a group of services linked by one or more operation paths. The reputation of the services or the service compositions can be derived from the reputation of the workflow in different ways. Intuitively, we adopt an equal-share model where the involved services (S), service associations ($S^{<>}$) and the service compositions ($S^{\{\}}$) share the reputation of the workflow equally

$$\exists S \in W_i \quad r^{Wi}_{S} = r^{Wi}_{Di}, \tag{6}$$

$$\exists S^{<>} \in W_i \quad r^{Wi}_{S<>} = r^{Wi}_{Di}, \tag{7}$$

$$\exists S^{\{\}} \in W_i \quad r^{Wi}_{S\{\}} = r^{Wi}_{Di}. \tag{8}$$

The assumption underpinning the equal-share model is that if the groups of users are aware of the workflow, they are implicitly aware of the services contained within. In other words, the services, service compositions and the workflow are equally reputable.

A more sophisticated approach is to adopt a *fair-share* model. The *fair-share* model considers the factor that different components of the system offer different levels of contributions towards the systems performance. For instance, if a workflow simply has one activity which is invoking a service, the service provides the entire functionality of this workflow. If a workflow contains a service composition of two services, their composition provides the whole functionality while each of them provides only a portion of that functionality. Under this intuition, the *fair-share* model distributes the reputation of the workflow to the services and service compositions according to their contributions to the workflow. The importance of each element in the workflow (including services and compositions) will first need to be assigned

$$V^{Wi} = v_1, v_2, v_3, \ldots, v_n, \tag{9}$$

where W_i is a workflow containing n elements and $v \in [0, 1]$ and it is designed by D_j. Then the reputation of individual elements are computed as follows:

$$\forall S_a \in W_i \quad r^{Wi}_{sa} = v_a r^{Wi}_{Dj} \tag{10}$$

$$\forall S^{<>}_b \in W_i \quad r^{Wi}_{Sb<>} = v_b r^{Wi}_{Dj} \tag{11}$$

$$\forall S^{\{\}}_c \in W_i \quad r^{Wi}_{Sc\{\}} = v_c r^{Wi}_{Dj}. \tag{12}$$

The importance can be derived in multiple ways depending on the implementation details. One could, for example, refer to the popularities of other workflows containing the element; or for simplicity, assume equal importance for all services and calculate importance according to the number of services involved, i.e., in a composition of two services, each contributes half of the functionality. More details about *fair-share* can be found in [47].

A particular service or a service composition may be used in multiple workflows designed. The instances of their use all contribute to their reputations in the community, i.e., the more times a service is used in different workflows, the more reputable this service becomes. An aggregate reputation of the service or the service composition can be calculated according to the individual reputations derived from the workflows that incorporate them. This aggregate reputation suggests the reputation of the service or the service composition within this community. The aggregate reputation of a given service S_i (R_{Si}), a given association $S^{<>}_i$ ($R_{Si<>}$) and the aggregate reputation of a given service composition $S^{\{\}}_i$ ($R_{Si\{\}}$) are computed as follows:

$$R_{Si} = f_{Rs}\big(r^{W1}_{Si}, r^{W2}_{Si}, \ldots, r^{Wn}_{Si}\big) \tag{13}$$

$$R_{Si<>} = f_{Ra}\big(r^{W1}_{Si<>}, r^{W2}_{Si<>}, \ldots, r^{Wn}_{Si<>}\big) \tag{14}$$

$$R_{Si\{\}} = f_{Rc}\big(r^{W1}_{Si\{\}}, r^{W2}_{Si\{\}}, \ldots, r^{Wn}_{Si\{\}}\big). \tag{15}$$

Individual reputation scores are computed based on the reputation of the workflow designer and the popularity of the workflow. Thus, a service or service composition with a high aggregate reputation entails that: i) it has been involved in the workflows that are used extensively by the users; or ii) it has been involved in the workflows that are designed by very reputable designers; or iii) both i) and ii).

3.2 Reputation Bootstrapping

It is possible that for a workflow, one or both of the elements in the function in (1) are missing. For instance, newly posted workflows will have no popularity. In these cases, we need to bootstrap the workflow reputation.

For newly posted workflows without popularity $p^{Wi} = \phi$, the calculation remains the same, as we can calculate the reputation of the workflow solely based on the reputation of the author

$$r^{Wi}_{Di} = f_r(R_{Di}, \phi) = f_r(R_{Di}). \tag{16}$$

In this case, this workflow simply 'inherits' the reputation of the designer, which is an estimation based on his/her past contributions. Alternatively, to promote potentially valuable workflows, we can consider applying an initial popularity to bootstrap the reputation

$$r_{Di}^{Wi} = f_r(R_{Di}, \phi) = f_r(R_{Di}, p^{\#Wi}), \tag{17}$$

where $p^{\#Wi}$ is the adjusted popularity

$$p^{\#Wi} = p^{Wi} + p_{Di}^{Bp} \tag{18}$$

with p_{Di}^{Bp} being the average popularity of the other workflows defined by D_i. Given the set $\{W\}^{Di}$ containing all workflows defined by D_i

$$p_{Di}^{Bp} = \frac{1}{N}\sum_{j=1}^{N} p_{Di}^{Wj}, \quad W_j \in \{W\}^{Di} \; and \; W_j \neq W_i. \tag{19}$$

The intuition is that, if one person does not show sharp improvement or deterioration in his/her profession, we can assume that the popularity of a new workflow defined by the person is likely to be the same as those defined in the past.

According to our modeling method, it is also recognized that new myExperiment members will be put into a disadvantageous position if assigned zero reputation upon joining. In order to encourage new users to contribute to the community, this disadvantage should be minimized if cannot be eliminated. Bootstrapping ones reputation refers to assigning or adding an initial value to a newcomers reputation [25], so that his/her newly designed workflows are reasonably competitive among the workflows designed by others.

Such an objective can be achieved in multiple ways. For example, we can initialize one's reputation based on the reputations of his/her connections. Similarly, if the community provider can incorporate endorsement techniques [24], newcomers can present the credentials of any existing users who are willing to endorse them. However, these approaches will likely to impose further assumptions on the community. An alternative, and a more generic approach can be taken by initializing one's reputation with the average reputation of all users. It is demonstrated in [11] that such an averaging technique provides the best results in terms of fairness and accuracy. Taking this as a starting position, we can adopt an averaging model in this case. The bootstrapping reputation of a newly joined user R_{Di}^{Bp} can be calculated as follows:

$$R_{Di}^{Bp} = \frac{1}{N}\sum_{j=1}^{N} R_{Dj} \quad for \; all \; j \neq i, \tag{20}$$

where N is the total number of the users in the community. The newcomer's reputation will be boosted by adding this average reputation to the actual reputation computed in (5)

$$R_{Di}^{\#} = R_{Di} + R_{Di}^{Bp}, \tag{21}$$

where $R_{Di}^{\#}$ is the adjusted reputation for the newcomer D_i. For the reputation of the very first workflow designed by D_i, its popularity can be neglected (13) or bootstrapped (14) with p_{Di}^{Bp} being the average popularity of all the workflows in the repository

$$p_{Di}^{Bp} = \frac{1}{N}\sum_{j=1}^{N} p^{Wj}, \quad for \; all \; W_j \neq W_i. \tag{22}$$

In this way, a new workflow defined by newly joined users will obtain a fair treatment. With average designer reputation and workflow popularity, according to (2) the new workflow will obtain an average reputation. The bootstrapping reputation offers an advantage to the newcomers to offset their disadvantage inherent in not having any prior contributions. Clearly, this advantage should not be offered forever. This issue will be discussed in the next section when we introduce the *temporal sensitivity*.

3.3 Reputation Fading

An important aspect in reputation assessment is the *temporal sensitivity*, as reputation information of a service decays with time [38]. Given two workflows with the same popularity, the workflow that is designed earlier should have a lower reputation, as it suggests that the newer workflow has gained equal popularity more rapidly and is likely more up to date. Similar idea can be applied to the reputation of the authors.

Along those lines, the advantage offered to the newcomers should diminish in time so that eventually only their true reputations are counted (equation (3)). Thus, a bootstrapping fading factor $\alpha_{fd}^{Bp} \in [0, 1]$ must be incorporated into equation (10)

$$R_{Di}^{\#} = R_{Di} + \alpha_{fd}^{Bp} R_{Di}^{Bp} \tag{23}$$

$$\alpha_{fd}^{Bp} = f(\Delta t), \tag{24}$$

where Δt is the time difference between the present and the time of the event, i.e., when the newcomer joined. A linear or exponential function can be applied to reduce α_{fd}^{Bp} from 1 eventually to 0.

The significance of popularity of a workflow fades in time, i.e., a workflow can gain a high popularity if it is downloaded or used extensively. However, this popularity should decline if such a trend discontinues. Given that the statistics of its usage usually cannot decrease (e.g., the number of downloads), two popularity fading factors α_{fd}^{Pop} and β_{fd}^{Pop} should be incorporated into the model. The first factor α_{fd}^{Pop} is to diminish the bootstrapping popularity p_{Di}^{Bp} in time, while the second factor constantly reduces the popularity as the time passes. The two factors need to be computed using different functions, as α_{fd}^{Pop} is a multiplier, while β_{fd}^{Pop} is a subtrahend. Thus, the adjusted popularity $p^{\#Wi}$ then becomes

$$p^{\#Wi} = p^{Wi} + \alpha_{fd}^{Pop} p_{Di}^{Bp} - \beta_{fd}^{Pop} \tag{25}$$

$$\alpha_{fd}^{Pop} = f_1(\Delta t, z) \quad \beta_{fd}^{Pop} = f_2(\Delta t, z), \tag{26}$$

where Δt is the time interval difference between the present time and the time when this workflow is designed, and z is a catchall variable referring to other factors that may be considered.

Incorporating both the reputation bootstrapping and the reputation fading mechanisms into our earlier modeling, the reputation of a workflow will then be computed using both the adjusted designers reputation $R^{\#}_{Di}$ and adjusted workflow popularity $P^{\#Wi}$

$$r^{Wi}_{Di} = f_r\big(R^{\#}_{Di}, P^{\#Wi}\big). \tag{27}$$

Note that the popularity fading mechanism affects both the reputation of the workflow (2) and the reputation of the designer (according to (4) and (5)). This allows the convergence of reputation to a very small value as time passes, if the workflow is no longer used or the workflow designer stops contributing.

4 SERVICE RECOMMENDATION BASED ON REPUTATION

As noted above, two key recommendation needs have been recognized in ServiceMap, namely: 1) recommend the services that have been used together—undirected association rules recommendation; and 2) recommend the paths that link one service to another—directed service composition recommendation. Starting from ServiceMap, we now elaborate our approach.

The association rules obtained can be used to suggest other relevant services usually used together by peers. Feedback from caBIG [10] users shows that such association rules are quite helpful in terms of introducing relevant services from a large set into their experiments. However, due to the limited number of frequently utilized sets, the number of association rules that can be discovered according to the number of occurrences is low. The reputation of a service, on the other hand, can be high even when it has only been used in one workflow, as it inherits the reputation of the workflow and hence of the author. Based on (7) and (14), we have developed an algorithm, as shown in Algorithm 1, to find, for a particular service S_i, the set of the associated service groups $\{S^{<>}_{Si}\}$ that have reputations exceeding the threshold R^{Thres}.

Algorithm 1. Choosing Reputable Service Associations

Input: a set of workflows $\{wf\}$, a service S_a, a threshold R^{Thres}
Output: a set of associated service groups $\{S^{<>}_{Sa}\}$
$\{S^{<>}_{Sa}\} \longleftarrow \phi,\ Rep \leftarrow 0$
foreach $S^{<>}_{Sj} \in \{S^{<>}\}$ **do**
  if $S \in S^{<>}_{Sj}$ and $S^{<>}_{Sj} \notin \{S^{<>}_{Si}\}$ **then**
    if recommend based on $R_{S<>}$ **then**
      Compute $R_{Sj<>}$ (14) using (27) $Rep \leftarrow R_{Sj<>}$
    end
    else if *recommend based on* R_{Di} **then**
      Compute $R_{Sj<>}$ (14) using $r^{Wi}_{Di} = f(R_{Di}, \phi) = f(R^{\#}_{Di})$
      $Rep \leftarrow R_{Sj<>}$
    end
    else if *recommend based on* $p_{Sj<>}$ **then**
      Compute $R_{Sj<>}$ (14) using $r^{Wi}_{Di} = f(\phi, p^{Wi}) = f(p^{Wi})$
      $Rep \leftarrow R_{Sj<>}$
    end
    if $Rep > R^{Thres}$ **then**
    $\{S^{<>}_{Si}\} \leftarrow S^{<>}_{Si}$
    end
  end
end

In the previous algorithm, recommendations can be made based on three factors, namely: the reputation value of the association $R_{S<>}$, the reputation value considering only the designers reputation R_{Di}, and the reputation value considering only the popularity of the workflows $P_{Sj<>}$. The three mechanisms will be evaluated and compared in the experiment section.

Service composition recommendation provides a cross-workflow search technique. Our experience working with the caBIG community shows that this feature can be quite useful for scientists to explore best practices. In our view, the reputation of a service composition would yield more insights for the users.

As illustrated, the service compositions share the reputations of the workflows. For the service compositions that are used in different workflows that can be merged, we can synthesize the reputation of the merged larger service composition by merging their individual reputations. For instance, workflow W_a contains service composition $S^{\{\}}_a : S_1 \rightarrow S_2$, whilst workflow W_b contains service composition $S^{\{\}}_b : S_2 \rightarrow S_3$. Merging these two compositions, we obtain a larger composition $S^{\{\}}_c : S_1 \rightarrow S_2 \rightarrow S_3$. Although this composition may have not been used by any workflows designed, its reputation can be synthesized by utilizing the reputation of the two individual service compositions involved. That is

$$\exists S^{\{\}}_c = S^{\{\}}_a \cup S^{\{\}}_b \quad R_{Sc\{\}} = f(R_{Sa\{\}}, R_{Sb\{\}}). \tag{28}$$

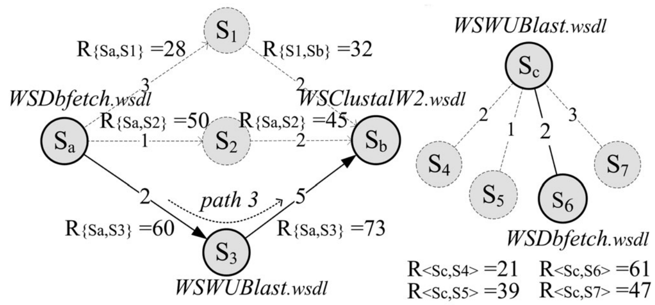
The above expression can be generalized to accommodate multiple service compositions.

A simple algorithm shown in Algorithm 2 can be followed to find the service composition that links two specific services with high reputation. Note that it relies on the ServiceMap to develop the directed operation network to identify all possible service compositions ($\{S^{\{\}}\}$) contained in the workflow set.

Algorithm 2. Choosing Reputable Service Compositions

Input: the set of all service compositions $\{S^{\{\}}\}$,S_a,S_b, a threshold R^{Thres}
Output: a set of reputable service compositions $\{S^{\{\}}_{Sa,Sb}\}$
$\{S^{\{\}}_{Sa,Sb}\} \leftarrow \phi, Rep \leftarrow 0$
foreach $S^{\{\}}_i \in \{S^{\{\}}\}$ **do**
  if $S_a \in S^{\{\}}_i$ and $S_b \in S^{\{\}}_i$ **then**
    Compute $R_{Si\{\}}$ using (15) and/or (28), apply similar mechanisms as in Algorithm 1 for the three types of recommendations: based on $R_{Si\{\}}$, R_{Di} and $p_{Sj\{\}} Rep \leftarrow R_{Si\{\}}$
    if $Rep > R^{Thres}$ **then**
      $\{S^{\{\}}_{Sa,Sb}\} \leftarrow S^{\{\}}_i$
    end
  end
end

To demonstrate the use of the recommendation algorithms, we have applied them to the example presented in Section 2 (Fig. 3). In Fig. 5, the service compositions and the associations in Fig. 3 are displayed together with the scaled



(a) Recommend reputable composition

(b) Recommend reputable association

Fig. 5. Examples of service recommendation.

reputations (linearly scaled to a value between 0 and 100, the latter being the most reputable) of the individual directed/undirected links. For simplicity, we apply summation for the elements in the functions. Same method is applied to other equations. As shown in Fig. 5, although the numbers of occurrences of the links are quite similar, their reputation scores are diverse. By using algorithms (1) and (2), we have selected the most reputable service associations (answer to Q1 in Section 2), and service compositions (answer to Q2 in Section 2). They are emphasized in the figure with thick solid lines, while other less reputable alternatives have been deemphasized with thin dashed lines. Clearly, the most reputable composition links services S_a and S_b through service S_3 (WSWUBlast.wsdl); whilst the most reputable association for service S_c is service S_6 (WSDbfetch.wsdl) among the alternatives.

5 Experiments

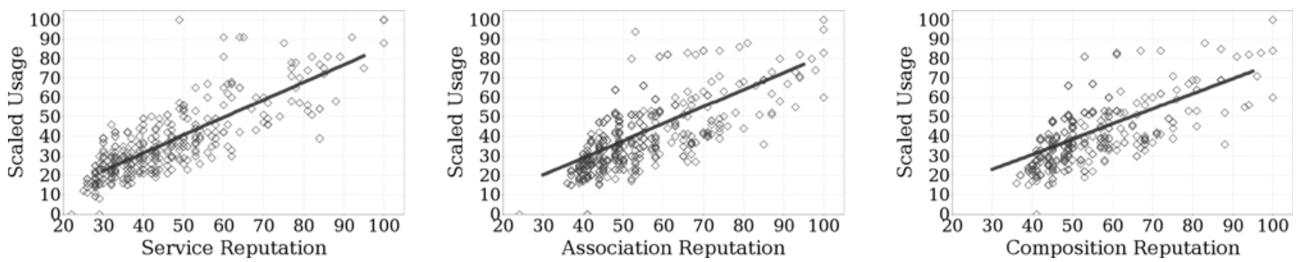
5.1 Methodology and Assumption

We have conducted a set of experiments using workflows on myExperiment to evaluate the effectiveness of ReputationNet. We first download all workflows on myExperiment to build the ServiceMap, which, as noted earlier consists of an undirected network of workflows-services and a directed network of service compositions. Next, we download the metadata pertaining to all workflows (including the ratings, view times and download times of each workflow) as well as the information related to each author (including his/her credit score, the number of friends and ratings) to construct the ReputationNet, which is superimposed on the ServiceMap. For simplicity, we use addition in the equations in Section 3 and 4 to calculate the reputation scores, that is, all the elements in the functions are added.

We assume that the download times of a workflow to a certain extent, reflect its usage. Certainly, it is possible that a workflow is downloaded many times but rarely used, but more likely a large download number indicates a lot of usages. We further assume that, in a social platform such as myExperiment, even without any recommendation mechanism, the community members will eventually find the high quality services or service compositions to use in their workflows, either through a long try-and-error process or through peers' suggestions. Based on these two assumptions, it follows that, high download times suggest high workflow usages which, in turn, suggest the high quality of the services, associated service groups and service compositions contained within. One may notice download times are used as an element in our modeling in Section 3 thus there is some natural correlation between the reputation scores and the download times. To address this shortcoming, we have conducted evaluations based only on the authors reputation (bootstrapping reputation) and cross-validation experiments to validate the legitimacy of the reputation scores (more details will be discussed later). Note that, although this assumption is based on reasonable arguments, the download times may not necessarily precisely describe the actual usage of the workflows and in turn, the accuracy of the reputation scores. We are merely using the download times as a yardstick to measure the different aspects of ReputationNet.

5.2 Results at a Glance

From myExperiment, we successfully downloaded and processed 1,557 workflows (the dataset is acquired in Feb 2013). Within downloaded workflows, we found 423 web services, 1,109 service association groups (association pairs) and 771 service compositions (two-service-composition). Such numbers have almost doubled since 2011 (as recorded in [42]). Five sets of experiment were conducted to evaluate various aspects of the ReputationNet and the observations are summarized in Figs. 6, 7, 8, 9, and 10. In the first three experiments (Figs. 6 and 8), we explored the correlation between the reputation of the service subjects (i.e., services, service associations and service compositions) and their performance (i.e., usage frequency) in different testing scenarios. The fourth experiment (Fig. 9) shows the effects of reputation bootstrapping and the fifth experiment (Fig. 10) examines the success rate the correlation between the reputation of service subjects and their chances of being used by designers in their workflows.

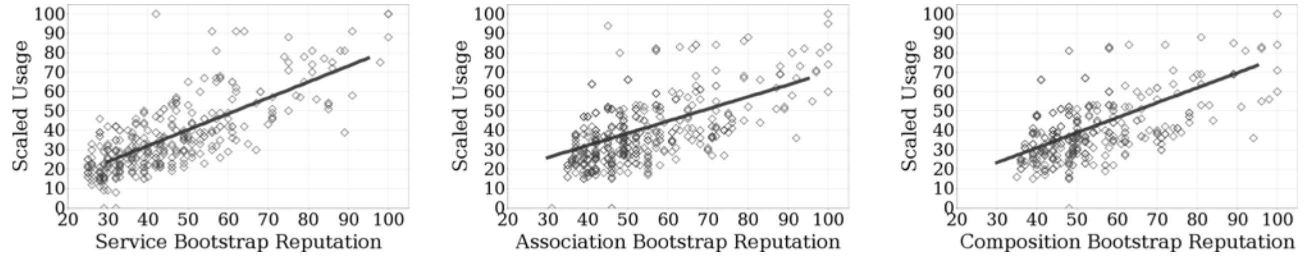


(a) Service reputation

(b) Association group reputation

(c) Service composition reputation

Fig. 6. Reputations of service subjects against their usage frequencies. The *x*-axis is the scaled reputation scores while the *y*-axis is the scaled usage. A linear regression line is draw for each of the sub-figures.



(a) Service reputation (b) Association group reputation (c) Service composition reputation

Fig. 7. Bootstrapping Reputations of service subjects against their usage frequencies. The *x*-axis is the scaled bootstrapping reputation scores while the *y*-axis is the scaled usage.

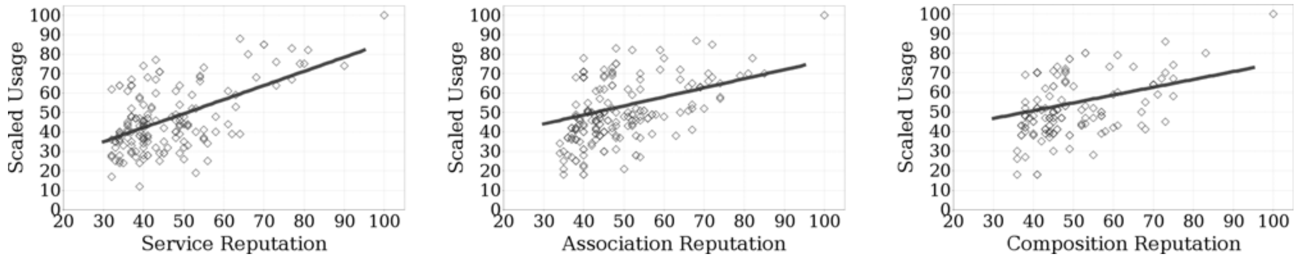
Fig. 6 shows the plots of the scaled (linearly scaled to a value between 0 and 100, the latter being the most reputable or used) reputations calculated using equations (13)-(15) and usages (number of downloads n_d) of the services (6a), service association groups (6b) and service compositions (6c). As mentioned earlier, the usages of the services/workflows indicate their performance in the community and therefore represent their quality. In this experiment, we aim to discover the correlation between the reputation scores and the actual performance of the services subjects. High correlation represents a close relation between the reputation scores and service/workflow performance and vice versa. We have evaluated the correlation by computing the Pearson correlation coefficient [36]. We also have studied the relative impact between the two variables through simple linear regression analysis. The purpose of correlation analysis is to measure and interpret the strength of a linear or nonlinear (e.g., exponential, polynomial, and logistic) relationship between two continuous variables. Pearson correlation coefficient (denoted as ρ) takes on values between -1 and +1, ranging from being perfectly negatively correlated (-1) to uncorrelated (0) and to perfectly positively correlated (+1). The purpose of simple regression analysis is to evaluate the relative impact of a variable (reputation) on the other (usage). In our experiments, we have applied the linear least squares method to obtain the regression lines.

We calculated the reputations of the workflows using (2), then derived the reputations of the contained services, associations and compositions using an equal-share model (6, 7, 8). The aggregate reputations are calculated using (13, 14, 15). Obviously, the distributions of their reputations and usages are not uniform, but the positive correlations between the usage and reputation are apparent. Fig. 6a plots the reputation scores against the usages of web services. It has a ρ of 0.82 (strongly positive), a regression slope of 0.9 with standard deviation 10.25. The regression slope indicates that for every unit increase in reputation, the value of its usage will increase on average 0.9. Fig. 6b plots the reputation scores against the usages of service association groups. It has a ρ of 0.72 (strongly positive), a regression slope of 0.87 with standard deviation 10.72. Fig. 6c plots the reputation scores of service compositions against their usages. It has a ρ of 0.95 (strongly positive) and a regression slope of 0.78 with standard deviation of 6.84. These figures suggest a strong positive relation between the calculated reputation and the observed usage frequency. In other word, the reputation scores computed can serve as a good indication of the performance of the services/workflows, i.e., reputable service subjects tend to have better performance.

To further verify this relation, in the second experiment, we excluded the popularity of the workflow p^{Wi} in (2) and computed the reputations solely using the reputations of the authors, that is, bootstrapping the reputations for all workflows (i.e., including the workflows with popularity $p^{Wi} \neq \phi$). In this approach, the effect of the download times on the reputations computed is greatly reduced, which are only considered when evaluating the contributions of the authors (4, 5). The results are plotted in Fig. 7. The plots in Fig. 7 are more scattered than those in Fig. 6, and the positive correlation remains high. Fig. 7a (web services) has a ρ of 0.76 (strongly positive) and a regression slope of 0.82 with standard deviation 11.53. Fig. 7b (service association groups) has a ρ of 0.61 (strongly positive) and a regression slope of 0.64 (strongly positive) with standard deviation 12.24. Fig. 7c (service compositions) has a ρ of 0.93 (strongly positive) and a regression slope of 0.77 with standard deviation of 7.84.

5.3 Cross Validation Experiments

The third experiment is a cross-validation test to evaluate the performance of the ReputationNet in a fabricated application scenario. The full set of workflows is partitioned into two subsets, i.e., the training set and the testing set. For workflows posted in each month, half of them are randomly selected into the training set while the remaining half goes to the testing set. This partition is to make sure the workflows in the training set and the testing set have similar timeline distributions. Next, we computed the reputations of the services, association groups and compositions using the workflows in the training set, and validate them with their usages observed in the workflows in the testing set. Through this approach, we are fabricating a real application scenario in which we investigate the relation between the reputations computed by ReputationNet and the usage frequencies of the service subjects (services, association groups and compositions) in the workflows that are not involved in the computation. The results are plotted in Fig. 8. Apparently, the data points we are able to collect are far less than those in Figs. 6 and 7. This is because the service subjects that have only appeared in either training set or testing set workflows will end up with zero usage or reputation and thus be excluded. We were only able to plot for 182 services, 346 service association groups and 218 service compositions, much less than half of the service subjects we have



(a) Service reputation (b) Association group reputation (c) Service composition reputation

Fig. 8. Cross-validation of reputations. The *x*-axis is the scaled reputation scores computed using the training set while the *y*-axis is the scaled usage frequencies computed using the testing set.

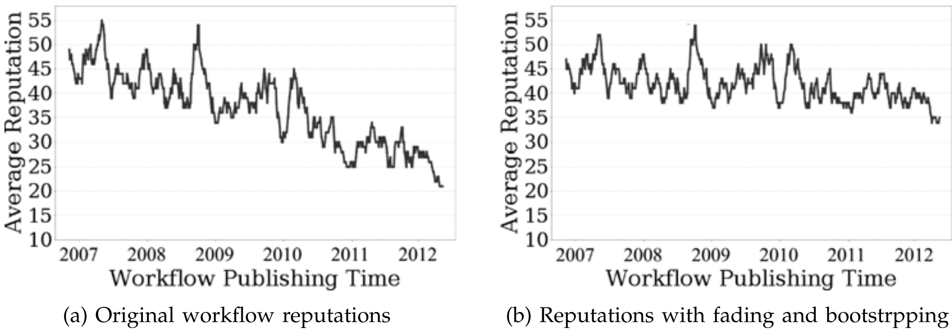
found in total. However, apart from the sparseness, there appears to be a clear positive correlation between the usage and reputation in the results. Fig. 8a has a ρ of 0.6 (strongly positive) and a regression slope of 0.7 with standard deviation 15.52. Fig. 8b has a ρ of 0.36 (moderately positive) and a regression slope of 0.45 with standard deviation 15.41. And Fig. 8c has a ρ of 0.34 and a regression slope of 0.36 with standard deviation of 15.06. Due to the limited number of samples that we are able to collect, we can see that the positive correlations in these figures are weaker than the previous ones and have larger deviations, but they are still evident.

5.4 Effects of Reputation Fading

In previous experiments, we have applied the reputation bootstrapping (21, 22) and fading (25, 26) mechanisms as elaborated in Section 3. Linear fading functions have been applied in (26), that is, the popularity of the workflows fades in time in a constant speed. The fading parameters in the functions have been assigned empirically based on our observations. In the real practice, more sophisticated fading algorithms shall be applied, and the fading parameters will need to be constantly tuned by domain experts. Here in Fig. 9 we demonstrate the effectiveness of reputation bootstrapping and fading. These two mechanisms work together to fade the reputation of the old workflows while bootstrapping the reputation of the new ones, therefore we are conducting this experiment to show their effects when both are applied. In Fig. 9a, we plot the unadjusted reputations of the workflows (range axis) against their publishing time (domain axis), then apply a 10 day moving average. We can observe from the plot that a downward trend is evident which indicates that, the workflows published recently will be put into a disadvantageous position in terms of reputations if the publishing time is not considered. Fig. 9b shows the same plotting after applying reputation bootstrapping and fading. It can be seen that the difference between the reputations of the newly posted workflows (e.g., after 2011) and the old workflows (e.g., before 2008) has been greatly reduced. This helps to promote the service subjects that have been freshly introduced to the community.

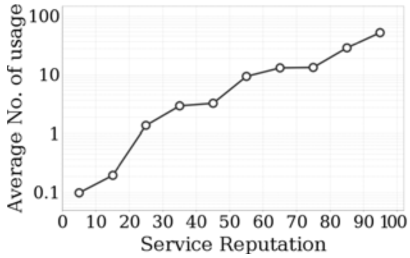
5.5 Success Rate Analysis

The last experiment conducted was to find out the success rate of the reputation scores. The success rate here is the rate or frequency at which a service is used by a domain scientist in his/her workflow. It is the correlation between the reputations of the service subjects, and the number of times they are chosen by the scientists. As explained earlier, we have assumed that the usage of a workflow directly reflects its quality. Under a similar intuition, if a service is indeed very useful for conducting a certain type of scientific experiment, naturally, we can expect the domain scientists to use the service in their workflows if the workflows are related to that experiment. In other word, we validate the reputations of the service subjects against their popularities among the workflow designers rather than users. If the reputation scores are reasonable, a service subject with a high reputation score should have a high success rate - it is favored by many scientists due to its high quality; and the opposite situation shall be applied to the ones with low reputation scores. Like the third experiment (Fig. 8), we partitioned the workflows into training set and testing set. Then we computed the reputations of the services subjects using the workflows in the training set, and counted the number of workflows in the testing set that uses them. In Fig. 10, we

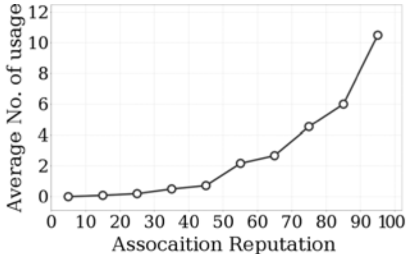


(a) Original workflow reputations (b) Reputations with fading and bootstrpping

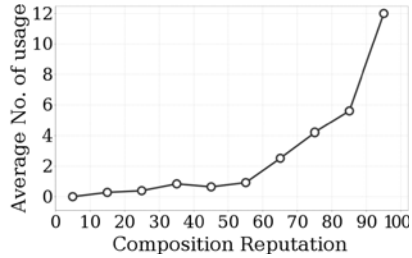
Fig. 9. Reputation fading and bootstrapping. The *x*-axis is the time at when the workflows are published while the *y*-axis is the average reputations of the workflows published.



(a) Success rate of services



(b) Success rate of association groups



(c) Success rate of service compositions

Fig. 10. Success rate experiment on service subjects. The *x*-axis is the scaled reputation scores of three service subjects, i.e., services, association groups and service compositions; the *y*-axis is the average number of the workflows that use these three subjects, respectively.

plot the reputations of the service subject against the number of workflows in which it has been used. The results are quite as expected, the higher the reputation of the service subject, the more workflows they are used within. Take services (Fig. 10a) for example, the services with low reputations (between 20 and 30) averagely have been used in 1.6 workflows while the ones with high reputations (greater than 90) averagely have been used in 55 workflows. The same pattern can be easily found in Figs. 10b and 10c.

5.6 Discussions

Through the experiments, we have observed that the reputations computed for the services, service association groups and service compositions are closely correlated to their usage frequencies. The higher the reputation, the more often the service subject is used in the community. We have also demonstrated the necessity of the reputation bootstrapping and fading mechanisms. Cross-validation experiments have been carried out to fabricate a real application scenario for ReputationNet. We have shown that the reputations of the service subjects computed using the training set; closely match the actual performance of them in the testing set. This indicates the legitimacy of the reputations produced by ReputationNet. Please note that, while we evaluated our modeling and algorithms with the scientific workflows on myExperiment, the general approach can be applied to other similar workflow or service composition platforms. The social aspects of such collaboration platforms are the true drive of the reputation based approach, regardless of the detailed implementations

6 Related Work

The concept of trust is not new. Trust has been studied in many disciplines including sociology [27], psychology [8], economics [19], and computer science [20]. Each of these disciplines has considered trust from different perspectives. There is no single consensus definition of trust in the literature. In general, trust is a measure of confidence that an entity or entities will behave in an expected manner. In this paper, we consider the reputation based trust [29], [30].

Reputation systems have benefitted electronic commerce in recent years. Amazon, eBay and Yahoo! Auction are examples of businesses that have deployed reputation systems successfully. These reputation systems use feedbacks from the consumers as the reputation measurement, and have received considerable attention in the literature [5]. SPORAS [45] is one such centralized reputation model that extends the above mentioned models with more sophisticated characteristics to model trust dynamics. We have adopted and extended this concept and used along with reputation networks to develop a reputation-based recommendation framework.

Recommender systems exploit both implicit and explicit data sources to generate recommended list. Explicit data sources include user profiles, articulated friend networks, or group memberships. Recommenders often exploit explicit friendships or linkages to generate recommendation lists, e.g., the Friend-Of-A-Friend (FOAF) algorithm [6], [16]. User profile data can be used to identify articles and other content believe to be relevant to users [23]. Implicit data sources have become more common for recommenders in social networks for several reasons. Firstly, typical social networks represent relationships in a binary manner, i.e., friend or stranger. However, research has shown that many types of relationship exist, and factors such as closeness and trust can better represent these relationships and facilitate better recommendations [13], [12].

Trust relationships between members have been used in recommender systems. Trust models for social networks can be classified into three groups: graph-based trust models, interaction-based trust models, and hybrid trust models. The graph-based trust models exploit the inherent structural properties of social graph. For example, [15] proposed a method of creating a trust network on the Semantic Web by extending the FOAF schema to allow users to indicate a level of trust for people they know. The interaction-based trust models exploit the interaction data such as engagement and popularity in the STrust model [32], [31]. [25] was among the first to attempt to quantify relationship strength from interaction data in social networks.

Interaction-based social trust models consider interactions in the community to compute trust, but ignore the characteristics of the social network structures. Hybrid models were developed to exploit the benefits of both interactions and social graph. [44] proposed such a social trust model for applications such as content distribution and micro-blogs. These models have been exploited by the recommender system to generate personalized recommendations by aggregating the opinions of other users in the trust network [26], [2], [18]. For example, [18] used a social graph approach to recommend a node in a social network using a similarity. Massa and Avesani [26] proposed a trust-based recommendation system to search for trustable users by exploiting trust propagation, whereas [2] proposed several recommendation models to provide factual information. For further details and other examples, we refer to [37].

Automatic service composition is another intensively investigated topic in services computing. Various techniques have been developed to discover relevant services and compose them in a proper sequence [9]. Aalst proposed a framework named TomTom4BPM [1] that adopts process mining technique for various purposes, such as comparing the actual process execution with pre-modeled ones and dynamically navigating during process exceptions. Some works also relate to process mining, such as deriving patterns from past usage data to predict the most likely next-step in building visualization pipelines [22], and case base reasoning in finding a similar workflow and using it to suggest the next component to be included in a workflow [7].

These approaches can yield good results when services have complete metadata (input/output, pre/post conditions, QoS, etc), such that the composition problem can be translated into a well formalized one such as optimization and AI-based planning. In reality, however, many services are widely used without much metadata. Meanwhile, online workflow repositories (such as myExperiment) allow scientists to share successful experimental routines that contain best practices to compose services. Based on this observation, we have designed a framework to derive the credibility of the authors and popularity of the services to develop a service reputation network, to provide recommendations based on empirical workflows. This work is developed based on our earlier works on ServiceMap [42], [28] and service reputation [24], [25].

7 Conclusion

Although the sharing of service-based capabilities opens a gateway to resource reuse, in practice, such reuse is very low. This paper presents the ReputationNet framework, as an extension to ServiceMap to employ trust and reputation mechanisms for service recommendations in building scientific workflows. We have developed a novel model to capture the trust and reputation aspects of scientific Web Services and workflows. Based on the model, we have designed heuristic algorithms to recommend reputable services, service associations and compositions. We also conducted extensive experiments using the real-world data from the workflows on the myExperiment repository. Our experiment results demonstrated a strong positive correlation (with Pearson coefficient 0.82) between the reputation of the service subjects (services, associations and compositions) and their actual usage by the myExperiment community. This result in turn confirms the validity of the reputation scores computed and the capability of our framework in terms of service recommendation. It is thus reasonable to conclude that ReputationNet is capable of offering justified assessment of reputations and therefore the quality of the services, association groups - answer to Q1 in Section 2; and compositions - answer to Q2 in Section 2. ReputationNet can hence assist scientists in choosing proper services to compose their scientific workflows.

In the future, we plan to explore other uses of this reputation mechanism, for example, to give suggestions to service providers by comparing the reputation of their services with competitors'. We also plan to integrate ReputationNet with a workflow repository having a social platform. In this way we can collect user activities and feedbacks to examine the effectiveness of ReputationNet, and enhance it with, e.g., collaborative filtering.

References

[1] W. M. P. Aalst, "Using process mining to generate accurate and interactive business process maps," in *Proc. Int. Business Inf. Syst. Workshops*, 2009, pp. 1–14.

[2] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz, "Trust-based recommendation systems: An axiomatic approach," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 199–208.

[3] J. Bhagat, F. Tanoh, E. Nzuobontane, T. Laurent, J. Orlowski, M. Roos, K. Wolstencroft, S. Aleksejevs, R. Stevens, S. Pettifer, R. Lopez, and C. A. Goble, "BioCatalogue: A universal catalogue of web services for the life sciences," *Nucleic Acids Res.*, vol. 38 no. (suppl 2), pp. W689–W694, 2010.

[4] The Role of Provenance Management in Accelerating the Rate of Astronomical Research, Ewa Deelman, G. Bruce Berriman. Proceedings of Science, International Astronomical Union Special Session 5, Rio de Janeiro, Brazil, Aug. 2009.

[5] D. Chang and F. Hussain, *Trust and Reputation for Service-Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence*. New York, NY, USA: Wiley, 2006.

[6] J. Chen, W. Geyer, D. Dugan, M. Muller, and I. Guy, "Make new friends, but keep the old: Recommending people on social networking sites," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 201–210.

[7] E. Chinthaka, J. Ekanayake, D. Leake, and B. Plale, "CBR based workflow composition assistant," in *Proc. World Congr. Services*, 2009, pp. 352–355.

[8] K. S. Cook, T. Yamagishi, C. Cheshire, R. Cooper, M. Matsuda, and R. Mashima, "Trust building via risk taking: A cross-societal experiment," *Soc. Psychology Quart.*, vol. 2, no. 68, pp. 121–142, 2005.

[9] S. Dustdar and W. Schreiner, "A survey on web services composition," *J. Web Grid Services*, vol. 1, no. 1, pp. 1–30, 2005.

[10] A. C. V. Eschenbach and K. Buetow, "Cancer informatics vision: caBIG," *Cancer Informatics*, vol. 2, pp. 22–24, 2006.

[11] M. Feldman and J. Chuang, "The evolution of cooperation under cheap pseudonyms," in *Proc. Conf. E-Commerce Technol.*, 2005, pp. 284–291.

[12] J. Freyne, S. Berkovsky, E. M. Daly, and W. Geyer, "Social networking feeds: Recommending items of interest," in *Proc. ACM Conf. Recommender Syst.*, 2010, pp. 277–280.

[13] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 211–220.

[14] C. A. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li, and D. D. Roure, "myExperiment: A repository and social network for the sharing of bioinformatics workflows," *Nucleic Acids Res.*, vol. 38, pp. W677–W682, 2010.

[15] J. Golbeck, B. Parsia, and J. Hendler, "Trust networks on the semantic web," in *Proc. Cooperative Intell. Agents VII*, 2003, pp. 238–249.

[16] I. Guy, I. Ronen, and A. Raviv, "Personalized activity streams: Sifting through the 'River of News'," in *Proc. ACM Conf. Recommender Syst.*, 2011, pp. 181–188.

[17] I. Guy, S. Ur, I. Ronen, A. Perer, and M. Jacovi, "Do you want to know?: Recommending strangers in the enterprise," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 2011, pp. 285–294.

[18] C. W. Hang, and M. Singh, "Trust-based recommendation based on graph similarity," presented at the Proc. AAMAS Workshop Trust Agent Soc., Toronto, Canada, 2010.

[19] F. Huang, "Building social trust: A human-capital approach," *J. Inst. Theor. Econ.*, vol. 163, no. 4, pp. 552–573, 2007.

[20] D. Hughes, G. Coulson, and J. Walkerdine, "Free riding on gnutella revisited: The bell tolls?," *IEEE Distrib. Syst. Online*, vol. 6, no. 6, p. 1, Jun. 2005.

[21] M. Kanehisa, S. Goto, M. Furumichi, M. Tanable, and M. Hirakawa, "Kegg for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Res.*, vol. 38, pp. 355–360, 2010.

[22] D. Koop, “VisComplete: Automating suggestions for visualization pipelines,” *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1691–1698, Nov./Dec. 2008.
[23] H. Liu and P. Maes, “Interestmap: Harvesting social network profiles for recommendations,” in *Proc. IUI Beyond Personalization*, 2005, pp. 54–59.
[24] Z. Malik and A. Bouguettaya, “Rater credibility assessment in web services interactions,” *World Wide Web*, vol. 12, no. 1, pp. 3–25, 2008.
[25] Z. Malik and A. Bouguettaya, “Reputation bootstrapping for trust establishment among web services,” *IEEE Internet Comput.*, vol. 13, no. 1, pp. 40–47, Jan./Feb. 2009.
[26] P. Massa and P. Avesani, “Trust-aware recommender systems,” in *Proc. ACM Conf. Recommender Syst.*, 2007, pp. 17–24.
[27] G. Mollering, “The nature of trust: From geog simmel to a theory of expectation,” *Interpretation Suspension. Sociol.*, vol. 35, pp. 403–420, 2002.
[28] J. Yao, W. Tan, S. Nepal, S. Chen, J. Zhang, D. Roure, and C. Goble. “ReputationNet: A reputation engine to enhance servicemap by recommending trusted services,” in *Proc. Int. Conf. Service Comput.*, 2012, pp. 454–461.
[29] J. Yao, S. Chen, C. Wang, D. Levy, and J. Zic, “Modelling collaborative services for business and QoS compliance,” in *Proc. Int. Conf. Web Services*, 2011, pp. 299–306.
[30] S. Nepal, Z. Malik, and A. Bouguettaya, “Reputation management for composite services in service-oriented systems,” *J. Web Services Res.*, vol. 8, no. 2, pp. 29–52, 2011.
[31] S. Nepal, C. Paris, S. K. Bista, and W. Sherchan, “A trust model based analysis of social networks,” *Trust Management Comput. Commun.*, vol. 1, no. 1, pp. 3–22, 2013.
[32] S. Nepal, W. Sherchan, and C. Paris, “Building trust communities using social trust,” in *Proc. 9th Int. Conf. Adv. User Modeling*, 2012, pp. 243–255.
[33] T. Oinn, M. Greenwood, M. Addis, M. N. Alpdemir, J. Ferris, K. Glover, C. Goble, A. Goderis, D. Hull, D. Marvin, P. Li, P. Lord, M. R. Pocock, M. Senger, R. Stevens, A. Wipat, and C. Wroe, “Taverna: Lessons in creating a workflow environment for the life sciences,” *Concurrency Comput.: Practice Experience*, vol. 18, no. 10, pp. 1067–1100, 2006.
[34] T. Oinn, P. Li, D. B. Kell, C. Goble, A. Goderis, M. Greenwood, D. Hull, R. Stevens, D. Turi, and J. Zhao, “Taverna/myGrid: Aligning a Workflow System with the life sciences community,” in *Workflows for e-Science: Scientific Workflows for Grids*. New York, NY, USA: Springer-Verlag, 2007, pp. 300–319.
[35] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, “Service-oriented computing: State of the art and research challenges,” *IEEE Comput.*, vol. 40, no. 11, pp. 38–45, Nov. 2007.
[36] J. L. Rodgers and W. A. Nicewander, “Thirteen ways to look at the correlation coefficient,” *Amer. Statist.*, vol. 42, no. 1, pp. 59–66, 1988.
[37] W. Sherchan, S. Nepal, and C. Paris, “A survey of trust in social networks,” *ACM Comput. Surv.*, vol. 45, no. 4, pp. 1–33, 2013.
[38] W. Sherchan, S. Nepal, and A. Bouguettaya, “A trust prediction model for service web,” in *Proc. Int. Conf. TrustCom*, 2011, pp. 258–265.
[39] W. Tan, R. Madduri, A. Nenadic, S. Soiland-Reyes, D. Sulakhe, I. Foster, and C. Goble, “caGrid workflow toolkit: A taverna based workflow tool for cancer grid,” *BMC Bioinformat.*, vol. 11, p. 542, 2010.
[40] W. Tan, J. Zhang, and I. Foster, “Network analysis of scientific workflows: A gateway to reuse,” *IEEE Comput.*, vol. 43, no. 9, pp. 54–61, Sep. 2010.
[41] W. Tan, and M. Zhou, *Business and Scientific Workflows: A Web Service-Oriented Approach*. New York, NY, USA: Wiley-IEEE Press, 2013.
[42] W. Tan, J. Zhang, R. Madduri, I. Foster, D. D. Roure, and C. Goble. “ServiceMap: Providing map and GPS assistance to service composition in bioinformatics,” in *Proc. IEEE Int. Conf. Services Comput.*, 2011, pp. 608–615.
[43] I. Taylor, M. Shields, I. Wang, and R. Philp, “Distributed p2p computing within triana: A galaxy visualization test case,” in *Proc. 17th IEEE Int. Parallel Distrib. Process. Symp.*, 2003, p. 16.1.
[44] S. Trifunovic, F. Legendre, and C. Anastasiades, “Social trust in opportunistic networks,” in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2010, pp. 1–6.
[45] G. Zacharia and P. Maes, “Trust management through reputation mechanisms,” *Appl. Artif. Intell.*, vol. 14, no. 9, pp. 881–908, 2000.
[46] J. Zhang, W. Tan, J. Alexander, I. Foster, and R. Madduri, “Recommend-as-you-go: A novel approach supporting services-oriented scientific workflow reuse,” in *Proc. IEEE Int. Conf. Services Comput.*, 2011, pp. 48–55.
[47] S. Nepal, Z. Malik, and A. Bouguettaya, “Reputation propagation in composite services,” in *Proc. Int. Conf. Web Services*, 2009, pp. 259–302.

Jinhui Yao received the BE and PhD degrees from the University of Sydney, Australia, in 2008 and 2013, respectively. He is currently a research scientist in Palo Alto Research Center (PARC). Since joining PARC, he has been engaged in a range of research projects aiming to facilitate agile business processes, dynamic service compositions, and leverage analytic workflows for big data. His research interests include cloud computing, service oriented architecture, business process management, and big data. He is actively contributing in computing research community through publications, journal edition/reviewing and conference services, including ICSOC, ICWS, SCC, etc.

Wei Tan received the BS and PhD degrees from the Department of Automation, Tsinghua University, China, in 2002 and 2008, respectively. He is currently a research staff member in IBM T. J. Watson Research Center. From 2008 to 2010, he was a researcher at Computation Institute, University of Chicago and Argonne National Laboratory. At that time, he was the technical lead of the caBIG workflow system. His research interests include big data, cloud computing, service-oriented architecture, business and scientific workflows, and Petri nets. He has published more than 50 journal and conference papers, and a monograph Business and Scientific Workflows: A Web Service-Oriented Approach (272 pages, Wiley-IEEE Press). He is an associate editor of *IEEE Transaction on Automation Science and Engineering*. He served in program committee of many conferences and cochaired several workshops. He received the Best Student Paper Award from IEEE ICWS 2014, Best Paper Award from IEEE SCC 2011, Pacesetter Award from Argonne National Laboratory (2010), and caBIG Teamwork Award from the National Institute of Health (2008). He is member of the ACM and a senior member of the IEEE.

Surya Nepal received the BE degree from the National Institute of Technology, Surat, India, the ME degree from the Asian Institute of Technology, Bangkok, Thailand, and the PhD degree from RMIT University, Australia. He is a principal research scientist at CSIRO Digital Productivity & Services Flagship. His main research interest include the development and implementation of technologies in the area of distributed systems and social networks, with a specific focus on security, privacy, and trust. At CSIRO, he undertook research in the area of multimedia databases, web services and service oriented architectures, social networks, security, privacy and trust in collaborative environment and cloud systems and big data. He has more than 150 publications to his credit. Many of his works are published in top international journals and conferences such as VLDB, ICDE, ICWS, SCC, CoopIS, ICSOC, *International Journals of Web Services Research*, *IEEE Transactions on Service Computing*, *ACM Computing Survey* and *ACM Transaction on Internet Technology*.

Shiping Chen is a principal research scientist at CSIRO Digital Productivity & Services (DPS). He also holds an adjunct associate professor title with the University of Sydney through teaching and supervising PhD/Master research students. He has been working on distributed systems for more than 20 years with focus on performance and security. He published more than 100 research papers on these research areas. He is actively involved in computing research community through publications, journal edition/reviewing and conference services, including WWW, EDOC, ICSOC, ICWS, SCC, etc. His current research interests include secure data storage & sharing and secure multi-party collaboration. He is a senior member of the IEEE.

Jia Zhang is an associate professor at Carnegie Mellon University-Silicon Valley. Her current research interests include services computing, with a focus on collaborative scientific workflow, Internet of Things, and cloudbased big data management. She has coauthored one textbook titled Services Computing and published more than 120 journal articles, book chapters, and conference papers. She is an associate editor of IEEE Transactions on Services Computing (TSC) and of International Journal on Web Services Research (JWSR), and editorinchief of International Journal of Services Computing (IJSC).

David De Roure received the PhD degree from the University of Southampton. He is a professor of e-Research at the University of Oxford and the director of the Oxford e-Research Centre. Focused on advancing digital scholarship, he works closely with multiple disciplines including social sciences (studying social machines), digital humanities (computational musicology), and computer science (large scale distributed systems and social computing). He was closely involved in the United Kingdom e-Science programme and held a national role from 2009-2013 as the United Kingdom National Strategic director for Digital Social Research. He is a fellow of the British Computer Society.

Carole Goble is a full professor of computer science at The University of Manchester, where she runs a large e-Science research group. Her interests include distributed computing, semantic technologies, scientific workflows, e-laboratories and collaborative sharing environments for scientific digital assets, new forms of scholarly communication, and scientific software engineering. For more than 25 years, she has been with scientists, including biomedical sciences, social science, astronomy, library science, and biodiversity, and has developed widely used open source tools and infrastructure for data-intensive scientific tasks. She is a founding partner of the UK's Software Sustainability Institute.